Highlights

- Information from torso edges, not the central body, drives self-estimates of body size in women
- Information extraction was independent of bubble size in the bubble masking task used
- Normal eye fixations up and down the central torso remained despite the bubble mask
- Eye movements and diagnostic regions for self-estimates of body size are not necessarily equivalent

Title: **The visual cues that drive the self-assessment of body size: dissociation between fixation patterns and the key areas of the body for accurate judgement**

Authors: Kamila R. Irvine[a], Kristofor McCarty[a], Thomas V. Pollet[a], Katri K. Cornelissen[a], Martin J. Tovée[b], Piers L. Cornelissen[a]

[a]Department of Psychology, Faculty of Health & Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK

[b]School of Psychology, College of Social Science, University of Lincoln, Lincolnshire, LN6 7TS, UK

CORRESPONDENCE TO:

Piers Cornelissen: piers.cornelissen@northumbria.ac.uk

Declarations of interest: none

Abstract

A modified version of the bubbles masking paradigm was used in three experiments to determine the key areas of the body that are used in self-estimates of body size. In this paradigm, parts of the stimuli are revealed by several randomly allocated Gaussian "windows" forcing judgements to be made based on this partial information. Over multiple trials, all potential cues are sampled, and the effectiveness of each window at predicting the judgement is determined. The modified bubbles strategy emphasises the distinction between central versus edge cues and localises the visual features used in judging one's own body size. In addition, eye-movements were measured in conjunction with the bubbles paradigm and the results mapped onto a common reference space. This shows that although observers fixate centrally on the torso, they are actually directing their visual attention to the edges of the torso to gauge body width as an index of body size. The central fixations are simply the most efficient way of positioning the eye to make this estimation. Inaccurate observers are less precise in their central fixations and do not evenly allocate their attention to both sides of the torso's edge, illustrating the importance of efficiently sampling the key information.

Key words: BMI, self-estimates, body size estimation, eye-movements, bubbles masking technique, visual cues.

1. Introduction

It is well established that people who suffer from anorexia nervosa or bulimia nervosa overestimate their own body size (e.g., Cornelissen, Johns, & Tovée 2013; Gardner & Bokenkamp, 1996; Probst, Vandereycken, Van Coppenolle, & Pieters, 1998; Slade & Russell, 1973; Tovée, Benson, Emery, Mason, & Cohen-Tovée, 2003), although the magnitude of this overestimation may also depend on a person's body mass index (BMI; Cornelissen et al., 2015, 2017). Body size overestimation is one of the most persistent of all the eating disorder symptoms, the severity of which predicts the long-term outcome of treatment (Fairburn, Cooper, & Shafran, 2003; Junne et al., 2019; Pike, 1998), and its persistence predicts the likelihood of relapse, which occurs at high rates (Berkman, Lohr, & Bulik, 2007; Castro, Gila, Puig, Rodriguez, & Toro, 2004; Channon & DeSilva, 1985; Herzog et al., 1999; Keel, Dorer, Franko, Jackson, & Herzog, 2005; Slade & Russell, 1973). It is therefore important that self-estimates of body size can be made accurately, that we understand how these judgements are made, how they may go awry, and to develop techniques to ameliorate this.

Two factors contribute to the estimation of one's own body size, both of which can be disturbed in eating disorders (Cash & Deagle, 1997): (1) an attitudinal component which captures the feelings that a person has about their body's size and shape, and (2) a perceptual component that has to do with the accuracy with which a person can judge the dimensions of their own physical appearance. Although more recent reviews exist, e.g., Skrzypek, Wehmeier, and Remschmidt (2001), they arrive at essentially the same conclusion. Measuring the attitudinal component of body image has proved to be relatively straightforward. Typically, psychometric tools are used to assess such attributes as body dissatisfaction and attitudes to body shape and weight (Evans & Dolan, 1993; Fairburn & Beglin, 1994). However, measuring the perceptual component of body size estimation has proved more challenging. A wide variety of methods have been tried, starting from image marking procedures (Askevold, 1975) and

moveable calliper techniques (Slade & Russell, 1973) to distorting photograph and video techniques (Gardner & Moncrieff, 1988; Probst, Vandereycken, & Van Coppenolle, 1995; Shafran & Fairburn, 2002). Most recently CGI (computer generated imagery) technology has been used to create standard stimuli or even personalized 3D avatars that accurately reflect BMI dependent body shape change (Cornelissen et al., 2015; Irvine et al., 2018; Mölbert et al., 2017; Szostak, 2018). In these perceptual body size estimation tasks, participants are typically presented images of either a standard model, or an avatar of themselves, usually on a PC monitor. The images vary in adiposity (indexed by BMI) and the participant's task is essentially to decide which image best corresponds to the body size they believe themselves to have. Our question is: what visual features do participants use to make these judgements about their own body size when they are viewing such stimuli?

1.1. Visual Cues to Body Size Judgements

Previous research suggests two potential sets of cues that may drive performance on perceptual body size estimation tasks: firstly, the width of the body in the stimuli and secondly, the cues within the body outline. The first set of cues are straightforward. Previous studies have noted that the width of the torso increases with increasing body mass index, particularly around the waist region (BMI) (e.g., Cornelissen et al., 2009a; Tovée et al., 1999; Tovée & Cornelissen, 2001). This "thickening" of the torso could thus provide an index of body mass. The second set of cues are internal to the body outline. These include the saliency of bony landmarks such as the collar bones or ribs, which become more obvious as body fat declines (George et al., 2011). Additionally, as the amount of body fat increases, it is deposited as rolls of fat, whose size and quantity could be used to estimate total body mass. Between these extremes, the pattern of texture gradients across the surface of the body can potentially provide a cue to the 3D shape of the body, such as size of the stomach (Cornelissen et al., 2013; Tovée et al., 2002).

In support of the first hypothesis, a principal component analysis (PCA) of images of female bodies varying in BMI, but facing forward in a standard pose, found that the change in torso width was described by principal component 1 (PC1), and this factor was the main predictor of body judgements (Tovée et al., 2002). Additionally, when the results of this PCA were used to create a set of artificial bodies, simply varying PC1 was sufficient to drive the perception of body weight change without varying any of the other shape dimensions (Smith et al., 2007a). This suggests that simple changes in torso width are sufficient to drive the perception of body mass.

This result is also consistent with a recent study which varied body orientation relative to the observer (Cornelissen et al., 2018). The observer had to discriminate between pairs of bodies in a 2-alternative forced choice task, based on differences in BMI. The finest discrimination occurred for the bodies presented either in profile or at 45^{o} relative to the observer, and the worst discriminations occurred when the bodies were presented in front-view. Most pertinently, the sensitivity of discrimination was predicted by the magnitude of the torso width change detectable by the observer. As BMI increases, the degree of change in torso width as a proportion of the total torso width, is greater in profile or at 45^{o} than in front-view. This is true for both CGI bodies and digital photographs of real bodies (Cornelissen et al., 2018). As a result, judgements in profile or at 45^{o} tend to be more accurate than those made in front-view. This difference in performance and its correlation with the saliency of the visual cues to change in torso width change, suggests that this is the cue that is being used to judge body size.

1.2. Eye-Movement Studies

Alternatively, there are also visual cues that are internal to the body outline that index overall body mass, and several studies suggest that in practice these are the cues being used. The evidence for this hypothesis is primarily based on eye-movement studies. For example,

women with anorexia nervosa fixate more on these body landmarks when making body size judgements than control observers and are significantly better than the control observers at judging the body size of low weight bodies (Cornelissen et al., 2015; George et al., 2012). This suggests that the use of these cues may form the basis of a successful strategy in judging lower BMI bodies. In addition, as mentioned above, increasing body fat changes the pattern of texture gradients and shading cues across the surface of the body within the body outline (Cornelissen et al., 2016b).

Several studies have suggested that stomach size, indexed through its depth, is a strong cue to BMI (e.g., Rilling et al., 2009; Smith et al., 2007; Tovée et al., 1999). Eye-movement data suggest control participants who are accurate at estimating their own BMI fixate primarily on the stomach. Critically, these fixations fall *within* the body outline (Cornelissen et al., 2009b, 2016b; George et al., 2012). This is true whether observers are judging bodies seen in front-view or viewed at a 45^{o} angle. If they were simply viewing the degree to which the stomach protrudes then their fixations should shift between central fixations on the torso in front-view to fixations on the edge body outline in the 45^{o} viewing angle. However, the fixations remain centrally located (Cornelissen et al., 2009b, 2016b; George et al., 2012). This is surprising, as if participants are asked to judge torso shape, they made eye-movements across the body and sequentially fixated on either side of the torso edge (Cornelissen et al., 2009b). This suggests that when viewing bodies at a 45^{o} angle, the optimal fixation strategy for estimating stomach depth would be to make fixations on both edges of the body corresponding to its outline. However, under these viewing conditions, observers whose fixations are not concentrated centrally within the body outline, and those who look more at the edge of the body are less accurate in their body mass judgements (Cornelissen et al., 2016b). Eye movement data like these therefore suggest that the principal cues being used to judge body mass are located within the body outline.

1.3. Dissociation Between Fixation Patterns and the Allocation of Attention

A potential key flaw with these eye-movement studies is the assumption that visual attention is always aligned directly with the line of sight. A number of studies have suggested that this may not necessarily be the case (e.g., Datta & DeYoe, 2009; Ehinger & Rosenholtz, 2016; Gegenfurtner, 2016). For example, in judgements of a basketball scenario, a contingent-gaze paradigm suggests that the position of the player with the ball is used as an "anchor point" for an observer's fixation while the relative position of the other players was estimated using the peripheral visual field (Ryu et al., 2013). Thus, a particular fixation point may just be a suitable point in the visual field from which to sample visual information using the retinal periphery, and not the complete focus of an observer's attention. This therefore raises an alternative account of the eye-tracking studies of body size estimation. It is possible that instead of extracting information *within* the body outline, the eye-movement pattern is actually an efficient foraging strategy which allows a wider attentional window to extract *edge-based* cues from the torso while using a central looking strategy.

It is well known that resolution acuity (i.e., the smallest separation between two points that allows them to be perceived as separate) drops off dramatically from the central fovea towards the parafovea and beyond (Anderson, Mullen, & Hess, 1991; Carrasco, 2011; Pelli & Tillman, 2008). This necessarily means that the apparent sharpness of the torso edges when sampled by a strategy of viewing the centre of the body, would be reduced; put simply, the torso edges would appear blurry. However, it is important to remember that the visual system's ability to resolve edge alignment, edge sharpness or smoothness, and curvature, i.e., exactly the kinds of low-level features that are likely to be needed to estimate the separation and shape of the torso edges, operate within the hyperacuity range (Carrasco, 2011). The phenomenon of hyperacuity is based not on the cone density of the retina, but on a cortical calculation which extrapolates from the limited sampling array to estimate a more detailed percept (Motter, 1998;

Gegenfurtner, 2016). This means that these spatial attributes can potentially be resolved to an accuracy often an order of magnitude finer than that of resolution acuity, even in the presence of a blurred stimulus. Therefore, there is no reason in principle why a foraging strategy that appears to blur the edges of the object being judged will impair the visual system's ability to discriminate the locations and shapes of those edges in calculating body size.

1.4. The Bubbles Masking Technique

A potential way of disambiguating these two possibilities, i.e., edge versus central image information and gauging the location of the attentional window during the perceptual judgement of body size, is the bubbles masking technique (Gosselin & Schyns, 2001). This technique is a psychophysics paradigm that has been used to determine which visual cues are being used in a categorisation task; i.e., which areas are diagnostic for a given judgement. For example, the technique has been used to reveal which facial features drive the distinction between neutral versus happy faces and male versus female faces. In the bubbles masking task, parts of the stimuli are revealed by randomly allocated Gaussian "windows." These are circular holes with blurred edges that perforate a uniform gray surface that overlies the stimulus (see Figure 1 for an illustration). On each trial, observers make a categorical judgement based on this partial information, e.g., "this is a male face" or, as in the current study, "that body is larger than mine." Over multiple trials, all areas in the stimulus image are sampled and from this unbiased sampling strategy, it is possible to calculate how effective each Gaussian window was at independently determining the behavioural performance (Humphreys et al., 2006). Thus, it should be possible to localise the areas of a body stimulus that are actually used when participants make self-estimates of body size.

But the bubbles masking technique has its own potential flaw. It is possible that the imposition of the bubble masks fundamentally changes the looking strategy (Gosselin &

Schyns, 2004; Murray & Gold, 2004). So, we address this problem by using an adapted bubbles strategy which emphasises the distinction between central versus edge featural information (Experiments 1 and 2). In addition, we also measure eye-movements to test whether the underlying search strategy, reflected in eye fixation patterns, has changed from the up and down the middle of the body fixation strategy reported by previous studies of self-estimation of body size (Experiment 3).

1.5. The Current Study

Here we ask what visual cues do participants use when judging their own body size? The literature reviewed above suggests that there are two potential sets of cues that participants could be using to make these judgements: (1) information about the separation of the torso edges and (2) information about body shape contained within the body outline. If the former case is true, we should expect to find a dissociation between where participants look on the stimulus bodies and the location of the regions on the bodies that are diagnostic for body size. Specifically, we predict that the eye fixations should lie along the vertical midline of the body stimuli, and the diagnostic regions should lie along the left and right torso edges. If, however, the latter case is true, both the diagnostic regions and the eye fixations should be spatially coincident, and both should be aligned with the vertical midline of the stimulus body.

In three experiments, we combine a modified bubbles masking technique together with eye movement recording to distinguish between these two possibilities. All the studies were completed by two sets of observers. In a pre-test screening process, we identified observers who were accurate at estimating their own body size, and observers who were inaccurate. By using both accurate and inaccurate observers we were able to compare the features important for an accurate judgement with the regions which lead to a misestimation. As discussed above, overestimation of body size in women with anorexia nervosa may arise from either one or both

of two factors; attitudinal or perceptual distortion. By testing nonclinical samples who overestimate body size compared to those who are accurate at estimating body size and who have the same psychological concerns, we can focus purely on perceptual factors as the basis of the overestimation. Ultimately, we intend to extend this research to compare diagnostic regions for self-estimates of body size in people with eating disorders with those from accurate and overestimating individuals without eating disorders. However, these experiments make heavy demands on participants. Therefore, as a first step in the introduction of the bubbles paradigm into this research area, we felt it appropriate to recruit participants who had no history of eating disorders.

1.6. Overall Experimental Strategy

In all three experiments, we recruited women with no history of eating disorders. For each experiment, we used a standard yes-no body size estimation task (described below) to identify a group of 12 women who estimated their body size accurately and a second group of 12 women who overestimated their body size. In addition, all participants were administered a standard battery of psychometric tasks to estimate their psychological attitudes regarding their body shape, weight, eating, and self-esteem, as well as report their symptoms of depression. This allowed us to ensure that, in each of the three experiments, the groups of accurate body size estimators and overestimators were comparable in terms of their psychological profiles, chronological age, and BMI. We then used the bubbles masking technique with large (Experiment 1) and small (Experiment 2) bubbles to identify the diagnostic regions that allowed participants to judge their own body size against the stimulus presented. On each trial of these tasks, participants had to decide whether the body in the masked image was smaller or larger than they believe themselves to be. Because the two groups of participants differed only in their accuracy at estimating their own body size (from the yes-no task), and did not differ in any other way, we could use a spatial analysis to compare the diagnostic regions for self-

estimates of body size between them. Finally, in Experiment 3, we ran an eye movement recording study to test whether the presence of the bubble masks caused a fundamental change in looking strategy in Experiments 1 and 2. Specifically, we needed to know whether participants had changed from an up-and-down the middle of the body viewing strategy, which we would expect to see in the absence of bubbles, to an alternative strategy in which they deliberately looked separately at the left and right torso edges.

2. Experiment 1

2.1. Method

The experimental procedures and methods for participant recruitment for this study were approved by the local ethics committee at Northumbria University.

2.1.1. Participants.

Pilot testing showed that the maxima and minima in the group differences in correctly responding in diagnostic areas that were biologically meaningful (e.g., edge of torso, central abdomen, and gap between thighs) could be detected using a sample size of between 4 and 11 participants per group (alpha = 0.05 and power = 80%). To offset attrition in participant numbers and/or unexpected sources of variability, we therefore recruited 12 participants per group.

To be eligible to take part in this study, participants had to be female (as assigned at birth), aged 18-35, with no history of eating disorders, and they had to have normal or corrected-to-normal vision. We recruited 41 females into Experiment 1 from staff and students at Northumbria University who carried out the initial psychometric and psychophysical tests. We defined body size overestimators as those whose point of subjective equality (PSE) from the yes-no-task (see below) was at least 2 BMI units above their measured BMI. Accurate body-size estimators recorded a PSE within +/-1 BMI unit of their measured BMI. According

to these criteria, we identified 12 accurate body size estimators and 12 overestimators from the initial sample of 41 consenting women and invited these individuals to complete the full study. The characteristics of these 24 participants are reported in Table 1.

2.1.2. Measures.

2.1.2.1. Psychometric and anthropometric measures.

To measure the attitudinal component of body image, participants completed a number of self-report questionnaires that measure body satisfaction and attitudes towards body shape, weight and eating.

2.1.2.1.1. Body Shape Questionnaire. The 16-item Body Shape questionnaire (BSQ-16) (Evans & Dolan, 1993) was used to assess participants' attitudes towards their body shape. Items are rated along a 6-point Likert-type scale ranging from *never* (scored as 1) to *always* (scored as 6). Items were summed to create a total score. A sample item is, "Have you been so worried about your shape that you have been feeling you ought to diet."

2.1.2.1.2. The Eating Disorders Examination Questionnaire. The Eating Disorders Examination Questionnaire (EDE-Q) is a 28-item self-report version of the Eating Disorder Examination (EDE) interview (Fairburn & Beglin, 1994). It contains four subscales: the Restraint subscale investigates the restrictive nature of eating, the Eating Concern subscale measures the preoccupation with food and social eating, the Shape Concern subscale measures dissatisfaction with body shape, and the Weight Concern subscale measures dissatisfaction with body weight. Participants report how many days out of the past four weeks they have experienced an item (e.g., "Have you been deliberately trying to limit the amount of food you eat to influence your shape or weight [whether or not you have succeeded]") on a 7-point Likert-type scale ranging from *No days* (scored as 0) to *Every day* (scored as 6). A global score of overall disordered eating behaviour and subscale scores were calculated by averaging the

appropriate items, and frequency data on key behavioural features of eating disorders is provided.

2.1.2.1.3. Beck Depression Inventory. The Beck Depression Inventory (BDI) was used to measure levels of depressive symptomatology (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). It is a behavioural checklist that contains 21 items. Each item is rated on a 4-point scale, ranging from 0 (no symptom of depression) to 3 (severe expression of depressive symptom). Items are summed.

2.1.2.1.4. Body mass index. BMI was calculated from their weight and height measured with a set of calibrated clinical SECA scales and a stadiometer, respectively.

2.1.2.2. Psychophysical measurements.

2.1.2.2.1. Yes-no task. In this study, we apply classical psychophysical methods (cf. Gardner, 1996) to measure two components of the participants' judgements of their own body size: (a) the point of subjective equality (PSE) and (b) the difference limen (DL). The PSE is the participant's subjective estimate of their body size. The DL is an estimate of how sensitive a participant is to changes in body size and equates to the smallest difference in body size that she can detect. To obtain these measurements, we use the method of constant stimuli in a yes-no forced choice paradigm. This allows a psychometric function to be estimated. Here, the psychometric function is a plot of the percentage of 'this image is larger than me responses' as a function of the BMI of the stimuli presented, and the curve tends to have a sigmoidal shape. The PSE is defined from the psychometric function as the BMI at which participants would respond 'larger than me' 50% of the time. The DL is the difference in the BMI of the stimuli falling between the 25% and 75% 'larger than me' response points (see Gescheider, 1997). This range captures the steepness of the psychometric curve. Participants who are very

sensitive to small differences in body size will have a steeper psychometric function with a correspondingly small DL.

In the yes-no task, participants were presented with a randomized sequence of images of a standard CGI female model, standing in three-quarter view (for details of stimulus image generation, see Cornelissen, 2016). Across the image set, BMI varied continuously from 12.5 to 44.5. On each trial of the task, one image was presented, and participants were required to decide whether the body depicted was larger or smaller than they believed themselves to be. Stimuli were presented on a 19" flat panel LCD screen (1280w ×1024h pixel native resolution, 32-bit colour depth) for as long as it took participants to make a decision. At the standard viewing distance of ~60cm, the image frame containing the female body subtended ~26° vertically and ~8° degrees horizontally. Each participant first judged seven images covering the whole BMI range (from 12.5 to 44.5 in equal BMI steps) presented in two separate blocks. Each stimulus image appeared 10 times in each block, and the order of presentation was randomized. Based on the responses from each block, the participants' point of subjective equality or PSE (the BMI they believe themselves to be) was calculated automatically by fitting a cumulative normal distribution. These two values were then averaged to give an initial estimate of the participant's PSE. Based on this initial estimate, the program presented a further set of 21 images (spread over a range of 5 BMI units centred on the participant's initial PSE, at a spacing of 0.25 units per image) for the participants to judge. Each image was presented ten times in randomized order. This final set of judgements allowed us to plot the full psychometric function (i.e. the percentage of 'larger than me' responses on the y-axis as a function of stimulus BMI on the x-axis) and use probit analysis off-line to calculate a definitive estimate of PSE as well as the difference limen or DL (that is how sensitive participants are to changes in BMI). Participants were classified as accurate at body size estimation if their PSE

was within +/1 BMI unit of their measured BMI and overestimators if their PSE was > 2 BMI units above their measured BMI.

2.1.2.2.2. Bubbles masking task. We built a bubbles masking task that was inspired by, but different from, the Bubbles paradigm developed by Gosselin and Schyns (2001). In these authors' task, like ours, on every trial, participants are given a partial view of a stimulus through a set of Gaussian windows (i.e., circular holes with blurred edges, see Figure 1). The holes are punched, as it were, through a gray overlay that covers the stimulus image. In Gosselin and Schyns (2001), the centre of any one Gaussian bubble can be located at any pixel location in the stimulus image. However, in the current study, we were asking whether information from the edges of the body outline, or the midline of the body, primarily drives decisions about self-estimates of body size. For this reason, we wanted to constrain the location of the mask bubbles into three columns. Bubbles in the left column of the stimulus overlay the right body edge and allowed participants to see this edge only. (Here we use the anatomical convention where left refers to the left side of the person in the stimulus image, from their point of view.) Bubbles in the middle column overlay the midline of the woman in the stimulus, thereby restricting participants' view to the midline of the body only. Bubbles in the right column of the stimulus overlay the left body edge, and restricted participants' view to that region only (see Figure 1a). This approach meant that we could carry out a spatial analysis of percentage correct responses at each fixed bubble location, and explicitly test for differences in body size classification between bubbles in the midline versus the two edge columns.

Bubbles were created dynamically as the program ran the task. On each trial, a stimulus image was covered by an opaque grey overlay (RGB: 64, 64, 64 on a 0-256 range), punctured by transparent bubbles whose centres were defined by the centres of an invisible, rectangular grid of squares 3(w) × 9(h), corresponding to the three columns (left edge, midline, and right edge). Each square of the grid measured 100 × 100 pixels. In Experiment 1, the transparency

of the bubbles followed a 2D Gaussian distribution with a standard deviation of 0.56 degrees. On each trial of the task, a subset of the bubble locations was chosen at random from this 3 × 9 array to be transparent, and participants had to decide, and respond by button press, whether the underlying image (drawn from the same stimulus set as the yes-no task) was larger or smaller than the participant believed themselves to be. Half of the images presented were larger, and half of them smaller, and the order of image presentation was randomized across trials. The particular pair of images presented to each participant were chosen based on their difference limen (DL) in the yes-no task. The smaller image corresponded to the 25% response rate in the yes-no task and the larger image the 75% response rate. Like Gosselin and Schyns (2001), we sought to maintain participants' performance in the bubble mask task at ~75% correct across the 2000 trials of the task. To do this, we calculated the correct response rate after every 20 trials, and reduced the bubble count by 1, kept it the same or increased it by 1 depending on whether the participant's responses were below, at or above criterion (within +/- 15 %).

2.1.3. Procedure. To maximize participant's vigilance and minimize their fatigue, they typically completed the experiment over the course of three sessions on three consecutive days. On the first day, in a quiet, private testing room, participants gave written consent to take part having read the study information sheet. Next, over the course of ~ 40 minutes, their height and weight were measured, they were asked to complete the psychometric questionnaires, and finally complete the yes-no psychophysical task. Participants who were eligible to complete the full study (i.e., they fit the criteria either for accurate or overestimation of body size) carried out the bubble masking task over the course of the next two sessions, each of which lasted about 60 minutes. Trials were presented back to back, each new trial triggered by the participant's button response. A pause was included after every 140 trials, giving the

opportunity for a break. Once all tasks were completed, participants were verbally debriefed and given the opportunity to ask questions about the study.

2.2. Results

2.2.1. Univariate statistics. The right-hand columns in Table 1 show the output of pairwise comparisons of the two group means, adjusted for multiple comparisons, using the bootstrap resampling method with 10,000 bootstrap samples in PROC MULTEST (SAS v9.4, SAS Institute, North Carolina, USA). The effect sizes (Cohen's *d*) for these comparisons, together with their 95% CI, are also included (Kadel & Kip, 2012). Despite some of the Cohen's *d* values representing medium-to-large effect sizes, almost all of them include 95% confidence intervals that include zero. This is likely attributable to the relatively small number of participants. The only confidence intervals that do not include zero, correspond to very large effect sizes, are these also associated with statistically significant pairwise comparisons. Table 1 confirms that accurate estimators were within ~0.25 BMI units of their actual BMI, on average, as compared to overestimators who overestimated by ~4 BMI units. With respect to the World Health Organization's BMI classification scheme (World Health Organization, 2003), the numbers of participants who fell into the underweight, normal, overweight, and obese categories for the accurate and overestimating groups, respectively, were: 0, 11, 1, 0, and 1, 9, 2, 0. The mean BSQ scores shown in Table 1 are consistent with mild concern with body shape (Evans & Dolan, 1993). The mean BDI scores for the accurate and overestimating groups are both consistent with the mild range. The EDEQ subscales in both groups were within 1*SD* of the normative means for women within this age group (Mond et al., 2006). Cronbach's alphas for the BDI, BSQ, and EDEQ in the two groups (combined) were .92, .95, and .94, respectively.

2.2.2. Where are the diagnostic regions for the accurate and overestimating groups? In Experiment 1, on each trial, the stimulus to be judged was visible through bubbles picked at random from an array of 3(w) × 9(h) bubble locations. By the end of the task, the number of times that any particular bubble location had been used, as well as the percentage of those presentations that were associated with a correct response were recorded for each participant. Therefore, a percentage correct could be calculated for every bubble location, separately for each participant.

The adaptive procedure ensured that participants' responses tracked close to the criterion we set for the masking task, namely that 75% of the choices they made across 2000 trials should be correct, and Table 1 confirms this. To achieve this criterion performance, both groups required on average a bubble count of ~5 (see Table 1). As Gosselin and Schyns (2001) argue, if all regions in our stimuli were equally informative about participants' perceptions of their own body size, then the percentage of correct responses at each location in our mask array should match the same criterion: i.e., the response rate for every bubble location should also be 75% correct. However, if there is a subset of areas in the stimuli that are particularly informative about the body size participants' believed they have, then we should expect the response rates in bubbles overlying these regions to be significantly higher than 75%. Such areas should correspond to regions that are diagnostic of participants' body size beliefs, according to the terminology of Gosselin and Schyns (2001). However, for this to be true, and for average performance across the set of trials to be 75% correct, we should also expect the response rates in bubble locations that overlie non-informative regions in the stimuli to be lower than 75% correct. Note that the non-informative regions do not necessarily need to be significantly lower than 75%. They might reach perhaps only ~72% for example, but nevertheless be widely distributed enough across the sample space so that the average across the whole space is 75%.

To test these predictions, we ran three generalized linear mixed models (GLMMs) of the normalized percentage responses across different bubble locations, using PROC MIXED in SAS v9.4 (SAS Institute, North Carolina, USA). To normalize the data, we calculated the mean percentage correct across all 3(w) × 9(h) bubble locations for each participant, and then subtracted these global means from the percentage correct for each individual bubble location, separately for each participant. For spatially sampled data, we cannot assume that the percentage correct responses at each bubble location are statistically independent of each other. Specifically, we must assume that percentage correct will covary across bubble locations, and that the magnitude of this spatial covariation is inversely proportional to the bubbles' proximity to each other. Therefore, in all three models we took account of the repeated measures within subjects – i.e., each subject was presented 27 mask locations in all (defined by row and column co-ordinates). In addition, we controlled for spatial covariance by incorporating the spatial variability into the statistical models by specifying a Gaussian spatial correlation model for the model residuals (Littell et al., 2006). The general form of the model we fitted was:

$$E[Y|u] = X\beta + Zu + e$$

Where E[Y|u] is the conditional probability of the outcome given the random model effects, Xβ are the fixed effects, Zu are the random effects, and e the error term. Spatial correlation was reflected in R, the covariance matrix of the model errors. The fixed effects in all models comprised two class variables: ROW (i.e., the index for each row of the grid of bubbles which could take values 1 to 9 inclusive) and COLUMN (i.e., the index for each column of the grid of bubbles which could take values 1 to 3 inclusive). This means that the location of each bubble in the 3 × 9 mask array was uniquely addressed, like an *x,y* coordinate, by the combination of the two fixed effect variables, ROW and COLUMN. Where relevant, we also included GROUP (i.e., accurate body size estimators versus overestimators) as a fixed

effect when we wanted to compare performance between accurate body size estimators versus overestimators. The most important outcomes from the statistical modelling were to identify:

MODEL 1: Where were the areas diagnostic of body size (i.e., > 75% correct) for accurate estimators?

MODEL 2: Where were the areas diagnostic of body size (i.e., > 75% correct) for overestimators?

MODEL 3: Where were the significant differences in diagnostic areas for body size comparing accurate estimators with overestimators?

To do this, for each model, we computed the predicted population margins from the GLMMs and compared them using tests for simple effects by partitioning the interaction effects, controlling for multiple comparisons. In other words, for MODELS 1 and 2, we used the fitted GLMMs to predict the percentage of correct responses in each bubble location and asked whether that percentage was significantly greater than 75%. These predictions are corrected for the repeated measures design, the spatial covariance in the data and the fact that we carried out multiple comparisons. For MODEL 3 we used the fitted GLMM to predict the difference in the percentage of correct responses comparing accurate body size estimators and overestimators, and asked whether each of these differences was significantly different from zero. An additional constraint for MODEL 3 was that a bubble location was only deemed to show a statistically significant difference between accurate and overestimators if that location had a response rate significantly greater than 75% ($p < .01$) from either MODEL 1 or MODEL 2, as well as showing a significant group difference. For completeness, we report the fixed effects in each model below, and then show the key outcomes, i.e., the predicted percentages of correct responses in each bubble location, in Figure 2.

The Type III tests of fixed effects for MODEL 1 were: ROW $F(4, 44) = 1.04$, $p = .40$; COLUMN $F(10, 110) = 25.02$, $p < .001$; ROW × COLUMN $F(40, 440) = 5.19$, $p < .001$.

The Type III tests of fixed effects for MODEL 2 were: ROW $F(4, 44) = 0.27$, $p = .90$; COLUMN $F(10, 110) = 12.98$, $p < .001$; ROW × COLUMN $F(40, 440) = 7.37$, $p < .001$.

The Type III tests of fixed effects for MODEL 3 were: GROUP $F(1, 22) = 0.00$, $p = .99$; ROW $F(4, 88) = 0.23$, $p = .92$, COLUMN $F(10, 220) = 36.91$, $p < .001$; GROUP × ROW $F(4, 88) = 1.16$, $p = .33$; COLUMN × GROUP $F(10, 220) = 2.39$, $p = 0.01$; ROW × COLUMN $F(40, 880) = 10.28$, $p < .001$; GROUP × ROW × COLUMN $F(40, 880) = 2.05$, $p < .001$.

In principle, a significant fixed effect of ROW means that, averaged across columns, there would be a significant linear increase/decrease in percentage correct responses as a function of ROW – i.e., a tilt to the 2D regression plane. Similarly, a significant fixed effect of COLUMN would mean that, averaged across rows, there would be a significant linear increase/decrease in percentage correct responses as a function of COLUMN. A significant interaction between ROW × COLUMN would mean that the degree of tilt in the 2D regression plane with respect to ROW, say, changes as a function of COLUMN. As the foregoing description of the fixed effects in the GLMMs makes clear, it is encouraging that we see statistically significant interactions between ROW and COLUMN in all three models. This strongly suggests that there are indeed statistically significant diagnostic regions of interest. However, analysis of the fixed effects alone cannot reveal the specific locations of the diagnostic bubbles. For this, we need post-hoc comparisons, to which we now turn.

The first two columns in Figure 2a show the outcomes of the analyses of simple effects from MODEL 1 and MODEL 2, for accurate body size estimators and overestimators, respectively. Circles correspond to mask locations where correct response rates were significantly higher than criterion (i.e., 75%), based on the GLMMs, and which can therefore

be considered diagnostic regions. The red/orange/yellow coloured overlay represents the averaged and smoothed raw data above criterion, referred to henceforth as a heat map.

For the accurate estimators, the circles *a* (80.4%, 95%CI 79.0 – 81.8%) and *c* (82.0%, 95%CI 80.7 – 83.4%) correspond to the peak LSmean response rates for the left and right columns of mask bubbles respectively, and circle *b* (78.4%, 95%CI 77.1 – 79.8%) is the closest mask bubble adjacent to both *a* and *c*. Circle *d* (78.4%, 95%CI 77.1 – 79.8%) corresponds to the peak LSmean response rate for the central column of mask bubbles. Therefore, while it is true that the central abdomen provides information that is diagnostic about body size for accurate estimators, the left and right torso edges appear to provide more information, and this difference is statistically significant for the left torso edge (i.e., the 95% confidence interval for *c* does not overlap with those for *b* or *d*).

For the overestimators, circles *e* (82.0%, 95%CI 80.8 – 83.3%) and *g* (80.2%, 95%CI 78.9 – 81.4%) correspond to the peak LSmean response rate for the left and right sides of the torso, and circle *f* (77.1%, 95%CI 75.9 – 78.4%) is the closest mask bubble adjacent to both *e* and *g*. Circle *h* (77.9%, 95%CI 76.7 – 79.2%) corresponds to the peak LSmean response rate for the central column of mask bubbles. Therefore, unlike the accurate estimators, the midline is providing diagnostic information about the face. As with the accurate estimators, the midline is also providing diagnostic information about the abdomen. However, the upper right torso and the left hip are providing more, and this difference is statistically significant for the upper right torso (i.e. the 95% confidence interval for circle *e* does not overlap with those for *f* or *h*).

The right most column in Figure 2a shows where diagnostic information about body size differs significantly between accurate and overestimators. Specifically, accurate estimators made significantly more use of information from the upper thigh gap and the left edge of the

abdomen (red/yellow colours), whereas overestimators made significantly more use of information from the right upper torso/arm and the face (blue/cyan colours).

2.3. Discussion

The results of Experiment 1 suggest that while both groups utilised information from the middle of the stimulus body as well as its edges, the edges provided the most diagnostic information (i.e., were more influential in driving participants' decisions in the categorisation task). Additionally, the two groups differed significantly in the edge cues used. While the accurate estimators made most use of the left flank and thigh gap, the overestimators used the face and right arm/chest area more. Interestingly, eye-tracking studies suggest that women with anorexia nervosa, who also overestimate body size, also fixate more on the face than nonclinical controls who accurately estimate body size (Cornelissen et al., 2016b). Accurate estimators also showed a distribution of diagnostic areas that are more evenly spread onto both sides of the body, whereas the diagnostic areas of overestimators showed a bias onto one side of the torso (see Figure 2a).

Even though the evidence from Experiment 1 suggests that body edges provide diagnostic information for body size judgements, some mid-body features were still used, i.e., the face and thigh gap. Therefore, in order to provide a more detailed picture of the edge cues used, we decreased the size of the bubbles from 100 × 100 pixels to 40 × 40 pixels in Experiment 2. With this strategy, by providing more bubbles that are smaller in size, a more detailed picture of the diagnostic information may be gathered.

3. Experiment 2

3.1. Method

3.1.1. Participants. The selection criteria and methods of participant recruitment were the same as for Experiment 1. Accordingly, we identified 12 accurate body size estimators and

12 overestimators from an initial sample of 41 consenting women, to take part in the complete study. These participants' characteristics are reported in Table 2.

3.1.2. Measures. The psychometric and psychophysical tasks were identical to Experiment 1. The only difference in the bubble mask task was that we used a finer scale rectangular grid of 9(w) × 21(h) squares (each of which measured 40 × 40 pixels), to locate the bubble centres. The transparency of these smaller bubbles followed a 2D Gaussian distribution with a standard deviation of 0.29 degrees, and the bubble count was increased or decreased by 2.

3.2. Results

3.2.1. Univariate statistics. Table 2 confirms that accurate estimators were within ~0.25 BMI units of their actual BMI, on average, as compared to overestimators who overestimated by ~4 BMI units. With respect to the World Health Organization's BMI classification scheme (WHO, 2003), the numbers of participants who were classified into the underweight, normal, overweight, and obese categories for the accurate and overestimating groups, respectively, were: 0, 10, 1, 1, and 2, 8, 2, 0. Cronbach's alphas for the BDI, BSQ, and EDEQ in the two groups (combined) were .92, .96, and .97, respectively. The mean BSQ scores shown in Table 2 are consistent with mild concern with body shape (Evans & Dolan, 1993). The mean BDI scores for the accurate and overestimating groups are consistent with the minimal and mild ranges respectively. The EDEQ subscales in both groups were within 1*SD* of the normative means for women within this age group (Mond et al., 2006). Table 2 shows that the adaptive procedure maintained participant performance very close to 75% correct in both groups, and that they required ~18-19 bubbles on average to achieve this performance.

3.2.2. Where are the diagnostic regions for the accurate and overestimating groups? The rationale for the analysis procedures in Experiment 2 was identical to those for

Experiment 1. Therefore, the treatment of data was the same, and we fitted the same 3 GLMMs as in Experiment 1. The only difference was in the resolution of the bubble mask, which comprised 9(w) × 21(h) bubble locations.

The Type III tests of fixed effects for MODEL 1 were: ROW $F(10, 110) = 10.44$, $p < .001$; COLUMN $F(22, 242) = 5.88$, $p < .001$; ROW × COLUMN $F(220, 2420) = 2.39$, $p < .001$.

The Type III tests of fixed effects for MODEL 2 were: ROW $F(10, 110) = 15.51$, $p < .001$; COLUMN $F(22, 242) = 8.21$, $p < .001$; ROW × COLUMN $F(220, 2420) = 3.13$, $p < .001$.

The Type III tests of fixed effects for MODEL 3 were: GROUP $F(1, 22) = 0.00$, $p = .99$; ROW $F(10, 220) = 24.7$, $p < .001$; GROUP × ROW $F(10, 220) = 1.21$, $p = .28$; COLUMN $F(22, 484) = 12.56$, $p < .001$; COLUMN × GROUP $F(22, 484) = 1.51$, $p = .06$; ROW × COLUMN $F(220, 4840) = 4.13$, $p < .001$; GROUP × ROW × COLUMN $F(220, 4840) = 1.38$, $p < .001$.

As before, the first two columns in Figure 2b show the outcomes from MODEL 1 and MODEL 2, for accurate body size estimators and overestimators respectively. Circles correspond to mask locations where correct response rates were significantly higher than criterion (i.e., 75%), based on the GLMMs. The heat maps represent the smoothed, averaged raw data above criterion. For the accurate estimators, the bubble locations corresponding to significant diagnostic information about body size are clustered continuously along the edge of the right lower chest and abdomen, the edge of the left waist and upper hip, and the thigh gap (again using anatomical conventions for left and right). The overestimators show a very similar pattern along the right edge of the upper body and a more extensive cluster along the left body edge extending to the chest. However, it appears that the overestimators do not make use of the

thigh gap. The right-hand column in Figure 2b shows where diagnostic information about body size differs significantly between accurate and overestimators. Specifically, accurate estimators made significantly more use of information from the upper thigh gap and a small region just to the right of midline in the upper abdomen (red/yellow colours). In comparison, the overestimators made more use of information on the right abdominal edge, as well as the left upper quadrant of the abdomen (blue/cyan colours).

3.3. Discussion

The results of Experiment 2 suggest that for both groups the edges of the body stimuli were instrumental in driving self-estimates of body size. Again, the two groups differed to some extent in cues used, with accurate estimators using the information about the thigh gap, and a region in the upper abdomen, while the overestimators used more cues from the right edge of the abdomen and an upper area of the abdomen. These results provide a more detailed picture of the diagnostic areas driving self-estimates of body size.

However, as described in the Introduction, it is possible that the presence of the bubbles and the partial view of the stimulus that this provides, changes the oberver's looking strategy. Therefore, we have measured the eye-movements of our participants to identify if the up-down looking pattern reported by prior studies of size estimation changes when a bubble mask task is used (Cornelissen et al., 2016b).

4. Experiment 3

4.1. Rationale

We wanted to know where participants were fixating when they carried out the bubble masking task with large and small bubbles. Therefore, in a third sample of participants, we recorded the movements of the right eye during 200 trials of each version of the bubble mask task. In addition, we also wanted to identify any differences in gaze patterns between the bubble

mask task as carried out in Experiments 1 and 2, compared to using the same size bubbles and the same task – i.e., judging whether the presented image was larger or smaller than the participant believed themselves to be, but now with all of the bubbles always set to transparent. These latter conditions, 200 trials with large bubbles all open and 200 trials with small bubbles all open, were the closest we could get to normal viewing using the bubbles task, and still permitting maximum visibility of all parts of the stimuli simultaneously, on every trial. Given that the view of the body per trial during the actual bubbles mask task is so restricted, we fully expected that there should be greater dispersion of fixations across space, when the data were binned over the course of 200 trials. Nevertheless, the critical question was whether participants adopted a different viewing strategy compared to what is usually seen when participants view non-masked bodies: i.e., looking up and down the midline of the body (see e.g., Cornelissen et al., 2016b). Specifically, given the evidence from Experiments 1 and 2 that the body edges provide diagnostic information for self-estimates of body size, we needed to know whether fixation patterns during the bubble masking task also split into two distinct distributions, with their peaks similarly aligned with the left and right body edges, instead of the midline.

4.2. Method

4.2.1. Participants. The selection criteria and methods of participant recruitment were the same as for Experiments 1 and 2. Accordingly, we identified 12 accurate body size estimators and 12 overestimators from an initial sample of 36 consenting women, to take part in the complete study. The characteristics of these 24 participants are reported in Table 3.

4.2.2. Measures. The psychometric and psychophysical tasks were identical to Experiments 1 and 2.

4.2.3. Eye movement recordings. Movements of the right eye were recorded with an Eyelink 1000 eye-tracker at a sample rate of 1000Hz. Stimuli were presented on a flat 19" CRT monitor while participants sat at a table with their heads restrained by a combined chin and

forehead rest. At the standard viewing distance of ~60cm, the image frame containing the female body subtended ~26° vertically and ~8° degrees horizontally. At the start of each block of 200 trials, participants' eye movements were calibrated using a 9-point calibration screen. Once the calibration procedure was validated, the experimental task began. We randomized the order of the four versions of the masking task: large bubbles, large bubbles open, small bubbles, and small bubbles open. While we did record participants' button responses in the task, there were not enough trials to warrant a spatial analysis of these behavioural data (i.e., $1/10^{th}$ of the number of trials in Experiments 1 and 2). Nevertheless, the average accuracy of responding over the 200 trials for large bubbles, large bubbles open, small bubbles, and small bubbles open was: 69%, 88%, 67%, and 87%, respectively, for accurate estimators. The equivalent performance for overestimators was: 69%, 98%, 69%, and 96%, respectively. Tests of location showed that all these values are significantly better than guessing (i.e., 50% accuracy), even though participants' performance had not stabilized at the ~75% criterion, which would be expected had they carried out all 2000 trials of the main tasks.

The Eyelink 1000 system uses a saccade-picker approach to identify saccades by applying an exclusive OR rule to three thresholds: velocity (30 degrees/sec), acceleration (8000 degrees/sec^2), and distance moved between samples (0.1 degrees). It then treats the rest of the (non-blink) data as fixations, assuming that the 'not in a saccade' condition is maintained for at least 50ms. The stated accuracy of the system is down to a resolution of 0.15°, though 0.25° to 0.5° is typical.

4.3. Results

4.3.1. Univariate statistics. Table 3 confirms that accurate estimators were within ~0.25 BMI units of their actual BMI, on average, as compared to overestimators who overestimated by ~4 BMI units. With respect to the World Health Organization's weight

classification scheme (WHO, 2003), the numbers of participants who fell into the underweight, normal, overweight, and obese categories for the accurate and overestimating groups, respectively, were: 0, 11, 0, 1, and 1, 9, 1, 1. Cronbach's alphas for the BDI, BSQ, and EDE-Q in the two groups were .90, .93, and .94, respectively. The mean BSQ scores shown in Table 3 are both consistent with mild concern with body shape (Evans & Dolan, 1993). The mean BDI scores for the accurate and overestimating groups are consistent with the minimal and mild ranges, respectively. The EDE-Q subscales in both groups all fall within 1*SD* of the normative means for women within this age group (Mond et al., 2006).

4.3.2. Where were participants fixating? The main question we wanted to address was whether participants were fixating primarily within the midline of the stimuli or along the body edges, during each of the four conditions: i.e., masking task with: large bubbles; large bubbles open; small bubbles; and small bubbles open. Therefore, our analyses focus on within task comparisons rather than between task comparisons. After blinks and saccades were removed from the eye movement time series, the only additional data filtering we applied was to remove the first 300msec post stimulus onset, as otherwise this would include the initial fixation which was determined by the fixation cross and not by the observer. In order to examine the spatial distributions of fixations, we constructed a sampling grid of square cells (20 × 20 pixels each) and applied it to the fixation data that were recorded within the central 600(w) × 1020(h) pixels of the stimulus array. This cell size (20 × 20 pixels) represents a compromise between capturing as many fixation samples per cell as possible to optimize statistical power (which ideally requires large cells) versus retaining good anatomical resolution (which ideally requires small cells) (cf. George et al., 2012). Having binned the fixation data in this way, we calculated the percentage of the total fixation samples in each bin, separately for each task and participant. These fixation density data were then converted to *z*-scores which are presented as heat maps in Figure 3.

Figure 3 shows clearly that, irrespective of whether they viewed stimuli through small or large bubble masks, or whether they were accurate body size estimators or overestimators, participants always showed a spatially more distributed gaze pattern during the bubble masking task as compared to viewing the stimuli when all bubbles were open. The critical question for the current study, however, is whether the gaze patterns for the bubbles task remain centred on the midline, or whether they break apart into two distributions: one centred on the left torso edge and the other on the right. Inspection of the black contours in Figure 3, which represent the three standard deviation limits in each heat map, would suggest that participants' fixations remained densest in the midline irrespective of task type or group assignment. To quantify this, we split each fixation density map into three columns of equal width, corresponding to the large bubble diameters at 100 pixels. We then calculated the total percentage of the fixation samples in each column, separately for each participant and for each task, and used PROC MIXED in SAS v9.4 (SAS Institute, North Carolina, USA) to test for differences between the average fixation density in each column. Table 4 shows the outcome including the post-hoc comparisons, controlled for multiple comparisons, between the left and middle columns and the right and middle columns of fixations. There is no case in Table 4 where both left and right columns of fixation data are significantly larger than the middle column. Therefore, we found no compelling evidence that participants' fixation patterns divided into separate distributions coincident with the edge regions diagnostic of body size. However, for accurate observers during the masking task, there was evidence that their gaze patterns shifted to the left, particularly in the chest region.

4.3.3. Direct comparison between eye fixations and psychophysical performance. Clearly, direct comparisons between Experiments 1 and 3 and between Experiments 2 and 3 were not feasible because the outcome measures, tasks, and participant groups were all different. Moreover, the spatial sampling of data in the three experiments was not directly

comparable. Nevertheless, we attempted to make approximate comparisons as follows. First, we resampled the eye movement data for each participant to match that for the bubble masking tasks. To do this, we used 20 × 20 pixel sample bins placed at the centres of the small and, separately, the large bubble masks. This procedure spatially co-registered the eye-movement data precisely with the large and small bubble mask psychophysical data. Then, we converted both the behavioural psychophysical data and the eye-movement data to *z*-scores, and re-ran the GLMMs, separately for the psychophysics and eye-movement data. This allowed us to compute marginal means (i.e., LSmeans in SAS) with their accompanying 95% confidence intervals for the data at each sample point, and these are plotted in Figure 4. In each case, the solid black lines represent the eye-movement data, and the solid white lines the psychophysical data. All error bars represent 95% confidence intervals in units of *z*-scores. The locations of the horizontal slices through the combined datasets are indicated by letter groups: *A*, *B*, & *C* and *D*, *E*, & *F*, for the large and small bubble mask datasets, respectively. Finally, there is a small horizontal offset in the x-axes for the eye-movement and psychophysical data, so that error bars do not overlap. Figure 4 confirms that eye fixations remained densest in the mid-line of the body, while the regions diagnostic of body size were concentrated on the edges.

5. General Discussion

In Experiment 1, the results of the modified bubbles technique (using the larger bubbles) suggest that the key areas of the image for accurate self-assessment of body size are on the edge of the torso at waist height on either side of the body. Both the left and right edges of the torso are of equal importance in making the judgement. Overestimating observers favour the right side of the image relative to the left side, as illustrated by the comparison of accurate and overestimators in Figure 2a. In Experiment 2, the results of the bubbles technique (using the smaller bubbles) suggests that the key areas are located along the outline of the torso on either side of the body and at the thigh gap. Once again, both sides of the body have equal

importance in accurate judgements, but there is a bias towards one side of the body in overestimators as illustrated by the comparison of accurate and overestimators in Figure 2b. It seems that an equal division of visual attention to both side of the torso outline may be key to accurate judgements.

A potential concern is that the use of the bubble masks significantly changes the looking strategy used to assess the stimuli (Gosselin & Schyns, 2004; Murray & Gold, 2004). However in face experiments, the diagnostic areas of the face identified by the bubbles techniques for a particular task are consistent with those identified using other methods, such as comparing the performance with isolated parts of the face (e.g., Bassili, 1979; Calder, Young, Keane, & Dean, 2000), using reverse correlation (Jack, Caldara, & Schyns, 2012; Yu et al., 2012), and eye-tracking (Blais et al., 2017). In Experiment 3, the addition of eye-tracking to the bubbles paradigm shows the visual fixations are clearly in the centre of the torso (Figures 3 & 4). This pattern of fixations is very similar to that reported by previous studies which have not used a masking paradigm, but have instead allowed a free, unoccluded view of the body stimuli during self-estimates of body size (Cornelissen et al., 2016b; George et al., 2012). This suggests that the use of the bubbles technique is not qualitatively altering the fixation pattern that our observers are using to estimate the size of their own body (Gosselin & Schyns, 2004). However, although the fixations fall within the centre of the stimulus torso, the key regions of the torso for accurate judgements are clearly on its edge (Figure 4). In short, the eye-movement results suggest a clear dissociation between fixation location and the location of the regions of the body stimuli that are diagnostic for self-estimates of body size.

At first, this dissociation might seem counterintuitive. The physical constraints of the retina mean that detailed spatial information can only be sampled from a small central area of around 2°, corresponding to the fovea (Levi, Klein, & Aitsebaomo, 1985). As a result, information in detail and colour can only be collected in small snapshots corresponding to an

observer's individual fixations (Miller & Bockisch, 1997). Thus, the failure to fixate the key regions of the body (as identified by the bubbles paradigm) so that the corresponding part of the image formed on the retina falls on the fovea is unexpected. Such a strategy should allow detailed analysis of the shape of these regions. Moreover, in a previous study in which participants were explicitly asked to judge torso shape (indexed by the waist-to-hip ratio), eye-tracking shows that fixations are initially made on one edge of the torso and then the participants' gaze moves across the torso to fixate the other edge (Cornelissen et al., 2009b). They do not make a simple central fixation as is seen here.

It is possible that the fixation on the centre of the torso may be serving as a convenient way to locate an image of the torso's left and right edges on the parafoveal region (the region of the retina surrounding the fovea). The parafoveal region supports a less detailed, lower resolution sampling than the fovea, but which is still sufficient to support the detection of the edges of the torso. This perception may be enhanced by the phenomenon of hyperacuity. In this perceptual process, the cortex extrapolates detail from the limited sampling of the parafoveal cone array and so is capable of finer discrimination than the retinal structure would suggest (Gegenfurtner, 2016; Motter & Belky, 1998; Ryu et al., 2013). So even though the centre of the torso is being fixated, information about the relative position of both torso edges can be derived from the periphery of the visual field and an estimate of the body width can be made. After all, just because the observer is fixating in the centre of the torso, that does not mean that her visual attention is focussed at the same position. Numerous studies have suggested that it is possible to direct attention at different parts of the visual field while at the same time fixating a separate part of the image (Evans et al., 2011; Motter & Belky, 1998), although it is unclear whether this allocation of attention across different parts of the visual field is achieved simultaneously or in rapid succession (Evans et al., 2011; Hutterman et al., 2013).

Thus, if one accepts that the width of the torso is a good cue to overall body mass, then the most efficient way of sampling the visual information that will allow you to make that judgement may not be to fixate on one edge of the torso and then move the eyes to fixate on the other edge of the torso. Instead, it may be quicker and simpler to foveate within the centre of the torso while directing your attention to the parafoveal regions of the retina corresponding to the edges of the torso. The previously reported difference in the pattern of eye-movements when estimating body size as opposed to judging body shape may be because although the parafovea can support enough spatial resolution to judge the relative position of the left and right torso edges (and so judge width), it may lack sufficient resolution to detect the subtler changes in the outline necessary to judge differences in torso shape (Cornelissen et al., 2009b).

This dissociation between the fixation pattern and the visual cues used in self-estimates of body size illustrates the danger of making assumptions based on eye-tracking data. Just because someone appears to look at a certain part of the body, it does not mean they are necessarily directing their visual attention to the same place. The assumption that these two visual activities are the same can lead to a misinterpretation of the data and mean that wrong conclusions are drawn on which body features are key to self-estimates of body size. In future research, it is important that eye-movement studies are paired with other techniques to localise which body features are used in a judgement, to either corroborate or clarify the results of the eye-tracking and avoid the wrong conclusions being made.

5.1. Clinical Implications

Given the dissociation between eye fixation and diagnostic regions we have found in this study of nonclinical women, it is clearly important to make the same measurements in women who have eating disorders. Based on an extensive review of the literature on visual processing in anorexia nervosa, Madsen, Bohon, and Feusner (2013) conclude that women with

anorexia nervosa struggle to process global features and tend to over-value local detail. Therefore, one possible outcome of applying the bubbles technique to a body size self-estimation task in anorexia nervosa might be to reveal a very non-specific, or diffuse pattern of diagnostic regions. On each trial, it is possible that participants might lock onto one or a very few bubbles to process only those local details. However, the particular bubble locations that they choose to focus on may be quite different from one trial to the next. When averaged over multiple trials, this could lead to widely dispersed and diffuse diagnostic regions. An alternative possibility might be that, in the face of such over-attention, women with anorexia nervosa may cling to a single well focused diagnostic region, say along just one body edge. If either of these outcomes were true, such findings might suggest new intervention strategies to retrain how sufferers attend to images of their body, thereby helping to prevent body size overestimation. We know that such an outcome could be useful, because recent perceptual training studies have shown clinically meaningful reductions in psychological concerns about body size, shape, and eating that last for up to a month post-intervention (Gledhill et al., 2016; Szostak, 2018).

5.2. Conclusion

In conclusion, the results of these studies using the modified bubbles technique suggest that the key visual cue used when making self-estimates of body size is the width of the torso, as judged from the relative position of the edges of the torso on either side of the body. Previous studies have found that the width of the torso increases with increasing BMI and so this would be a reliable cue to BMI status (e.g., Cornelissen et al. 2009a; Tovée & Cornelissen, 2001; Tovée et al., 1999). In the small bubbles condition, there is an additional important area of the image located at the position corresponding to the gap between the upper thighs. The diameter of the thighs is correlated with overall BMI (Ryan & Niklas, 1999) and so the "thigh gap" is a potential cue to overall adiposity, particularly for lower BMI bodies. The addition of eye-tracking to the paradigm suggests that observers use an efficient fixation strategy when

sampling the cues to body size, fixating centrally within the torso outline to estimate its width and thereby the BMI of the body.

Declarations of interest: There are no conflicts of interest

Funding: The first author was funded by a PhD studentship from Northumbria University.

References

Anderson, S. J., Mullen, K. T., & Hess, R. F. (1991). Human peripheral spatial resolution for achromatic and chromatic stimuli: Limits imposed by optical and retinal factors. *Journal of Physiology, 442*, 47-64. doi:10.1113/jphysiol.1991.sp018781.

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology, 37*, 2049–2058. doi:10.1037/0022-3514.37.11.2049

Beck, A. T., Ward., C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561-571. doi:10.1001/archpsyc.1961.01710120031004

Berkman, N. D., Lohr, K. N., & Bulik, C. M. (2007). Outcomes of eating disorders: A systematic review of the literature. *International Journal of Eating Disorders, 40*, 293–309. doi:10.1002/eat.20369

Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia, 50*, 2830–2838. doi:10.1016/j.neuropsychologia.2012.08.010

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 527–551. doi:10.1037/0096-1523.26.2.527

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*, 1484–1525. doi:10.1016/j.visres.2011.04.012

Castro, J., Gila, A., Puig, J., Rodriguez, S., & Toro, J. (2004). Predictors of rehospitalization after total weight recovery in adolescents with anorexia nervosa. *International Journal of Eating Disorders, 36*, 22–30. doi:10.1002/eat.20009

Channon, S., & De Silva, W. (1985). Psychological correlates of weight gain in patients with anorexia nervosa. *Journal of Psychiatric Research, 19*, 267-271. doi:10.1016/B978-0-08-032704-4.50031-0.

Cornelissen, K. (2016). *What does it mean to have distorted body image in anorexia nervosa?* Doctoral thesis, Northumbria University. http://nrl.northumbria.ac.uk/30330/

Cornelissen, K. K., Bester, A., Cairns, P., Tovée, M. J., & Cornelissen, P.L. (2015). The influence of personal BMI on body size estimations and sensitivity to body size change in anorexia spectrum disorders. *Body Image, 13*, 75-85. doi:10.1016/j.bodyim.2015.01.001

Cornelissen, K. K., Cornelissen, P. L., Hancock, P. J. B., & Tovée, M. J. (2016b). Fixation patterns, not clinical diagnosis, predict body size over-estimation in eating disordered women and healthy controls. *International Journal of Eating Disorders, 49*, 507–518. doi:10.1002/eat.22505

Cornelissen, K. K., Gledhill, L. J., Cornelissen, P. L., & Tovée, M. J. (2016a). Visual biases in judging body weight. *British Journal of Health Psychology, 21*, 555-569. doi:10.1111/bjhp.12185

Cornelissen, P. L., Cornelissen, K. K., Groves, V., McCarty, K., & Tovée, M. J. (2018). View-dependent accuracy in body mass judgements of female bodies. *Body Image, 24*, 116-123. doi:10.1016/j.bodyim.2017.12.007

Cornelissen, P. L., Hancock, P. J. B., Kiviniemi, V., George, H. R., & Tovée, M. J. (2009b). Patterns of eye movements when male and female observers judge female attractiveness, body fat and waist-to-hip ratio. *Evolution and Human Behavior, 30*, 417–428. doi:10.1016/j.evolhumbehav.2009.04.003.

Cornelissen, P. L., Johns, A., & Tovée, M. J. (2013). Body size over-estimation in women with anorexia nervosa is not qualitatively different from female controls. *Body Image, 10*, 103-111. doi:10.1016/j.bodyim.2012.09.003.

Cornelissen, P. L., Tovée, M. J., & Bateson, M. (2009a). Patterns of subcutaneous fat deposition and the relationship between body mass index and waist-to-hip ratio: Implications for models of physical attractiveness. *Journal of Theoretical Biology, 256*, 343-350. doi:10.1016/j.jtbi.2008.09.041

Datta, R., & DeYoe, E. A. (2009). I know where you are secretly attending! The topography of human visual attention revealed with fMRI. *Vision Research, 49*, 1037–1044. doi:10.1016/j.visres.2009.01.014

Ehinger, K. A., & Rosenholts, R. (2016). A general account of peripheral encoding also predicts scene perception performance. *Journal of Vision, 16*, 1-19. doi:10.1167/16.2.13

Evans, C., & Dolan, B. (1993). Body Shape Questionnaire: Derivation of shortened "alternate forms." *International Journal of Eating Disorders, 13*, 315-321. doi:10.1002/1098-108x(199304)13:3<315::aid-eat2260130310>3.0.co;2-3

Evans, K. K., Horowitz, T. S., Howe, P., Pedersini, R., Reijnen, E., & Pinto, Y., et al. (2011). Visual attention. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*, 503–514. doi:10.1002/wcs.127

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorder psychopathology: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*, 363-370. doi:10.1002/1098-108X(199412)16:4<363::AID-EAT2260160405>3.0.CO;2-#

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A "transdiagnostic" theory and treatment. *Behaviour Research and Therapy, 41*, 509–528. doi:10.1016/s0005-7967(02)00088-8

Gardner, R. M., & Bokenkamp, E. D. (1996). The role of sensory and nonsensory factors in body size estimations of eating disorder subjects. *Journal of Clinical Psychology, 52*, 3-15. doi:10.1002/(sici)1097-4679(199601)52:1<3::aid-jclp1>3.0.co;2-x

Gegenfurtner, K. R. (2016). The interaction between vision and eye movements. *Perception, 45*, 1333–1357. doi:10.1177/0301006616657097

George, H. R., Cornelissen, P. L., Hancock, P. J. B., Kiviniemi, V. V., & Tovée, M. J. (2011). Differences in eye-movement patterns between anorexic and control observers when judging body size and attractiveness. *British Journal of Psychology, 102*, 340-354. doi:10.1348/000712610X524291.

Gledhill, L. J., Cornelissen, K. K., Cornelissen, P. L., Penton-Voak, I. S., Munafò, M. R., & Tovée, M. J. (2016). An interactive training programme to treat body image disturbance. *British Journal of Health Psychology, 22*, 60–76. doi:10.1111/bjhp.12217

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41*, 2261–2271. doi:10.1016/s0042-6989(01)00097-9

Gosselin, F., & Schyns, P. G. (2004). No troubles with bubbles: A reply to Murray and Gold. *Vision Researc*h*, 44*, 471–477. doi:10.1016/j.visres.2003.10.007

Herzog, D. B., Dorer, D. J., Keel, P. K., Selwyn, S. E., Ekeblad, E. R., & Flores, A. T., . . . Keller, M. B. (1999). Recovery and relapse in anorexia nervosa and bulimia nervosa: A 7.5-year follow-up study. *Journal of the American Academy of Childhood and Adolescent Psychiatry, 38*, 829-837. doi:10.1097/00004583-199907000-00012

Hüttermann, S., Memmert, D., Simons, D. J., & Bock, O. (2013). Fixation strategy influences the ability to focus attention on two spatially separate objects. *PLoS ONE, 8*, e65673. doi:10.1371/journal.pone.0065673

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General, 141*, 19–25. doi:10.1037/a0023463

Junne, F., Wild, B., Resmark, G., Giel, K. E., Teufel, M., & Martus, . . . Zipfel, S. (2019). The importance of body image disturbances for the outcome of outpatient psychotherapy in patients with anorexia nervosa: Results of the ANTOP-study. *European Eating Disorders Review, 27*, 49-58. doi:10.1002/erv.2623

Kadel R. P., & Kip K. E. (2012, October). A SAS macro to compute effect size (Cohen's *d*) and its confidence interval from raw survey data. *Proceedings of the Annual Southeastern SAS Users Group Conference*, Durham, NC.

Keel, P. K., Dorer, D. J., Franko, D. L., Jackson, S. C., & Herzog, D. B. (2005). Postremission predictors of relapse in women with eating disorders. *The American Journal of Psychiatry, 162*, 2263-2268. doi:10.1176/appi.ajp.162.12.2263

Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Verner acuity, crowding and cortical magnification. *Vision Research, 25*, 963-977. doi:10.1016/0042-6989(85)90207-X

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberber, O. (2006). *SAS for mixed models* (2nd ed.). Cary, NC: SAS.

Madsen, S. K., Bohon, C., & Feusner, J. D. (2013) Visual processing in anorexia nervosa and body dysmorphic disorder: Similarities, differences, and future research directions. *Journal of Psychiatric Research, 47*, 1483-1491. doi:10.1016/j.jpsychires.2013.06.003

Miller, J. M., & Bockisch, C. (1997). Where are the things we see? *Nature, 386*, 550-551. doi:10.1038/386550a0

Mond, J. M., Hay, P. J., Rodgers, B., & Owen, C. (2006). Eating Disorder Examination Questionnaire (EDE-Q): Norms for young adult women. *Behaviour Research and Therapy, 44*, 53–62. doi:10.1016/j.brat.2004.12.003

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research, 38*, 1805–1815. doi:10.1016/s0042-6989(97)00349-0

Murray, R. F., & Gold, J. M. (2004). Troubles with bubbles. *Vision Research, 44*, 461-470. doi:10.1016/j.visres.2003.10.006

Pelli, D., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11*, 1129-1135. doi:10.1038/nn.2187

Pike, K. M. (1998). Long-term course of anorexia nervosa. *Clinical Psychology Review, 18*, 447–475. doi:10.1016/s0272-7358(98)00014-2

Probst, M., Vandereycken, W., Van Coppenolle, H., & Pieters, G. (1998). Body size estimation in anorexia nervosa patients: The significance of overestimation. *Journal of Psychosomatic Research, 44*, 451-456. doi:10.1016/s0022-3999(97)00270-5

Rilling, J. K., Kaufman T. L., Smith, E. O., Patel, R., & Worthman, C. M. (2009). Abdominal depth and waist circumference as influential determinants of human female attractiveness. *Evolution and Human Behavior, 30*, 21–31. doi:10.1016/j.evolhumbehav.2008.08.007

Ryan, A. S., & Nicklas, B. J. (1999). Age-related changes in fat deposition in mid-thigh muscle in women: Relationships with metabolic cardiovascular disease risk factors. *International Journal of Obesity, 23*, 126-132. doi:10.1038/sj.ijo.0800777

Ryu, D., Abernethy, B., Mann, D. L., Poolton, J. M., & Gorman, A. D. (2013). The role of central and peripheral vision in expert decision making. *Perception, 42*, 591-607. doi:10.1068/p7487

Slade, P., & Russell, G. (1973). Awareness of body dimensions in anorexia nervosa: Cross-sectional and longitudinal studies. *Psychological Medicine, 3*, 188-199. doi:10.1017/s0033291700048510

Smith, K. L., Tovée, M. J., Hancock, P. J., Bateson, M., Cox, M. A. A., & Cornelissen, P. L. (2007). An analysis of body shape attractiveness based on image statistics: Evidence for a dissociation between expressions of preference and shape discrimination. *Visual Cognition, 15*, 927-953. doi:10.1080/13506280601029515

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16*, 184–189. doi:10.1111/j.0956-7976.2005.00801.x

Szostak, N. M. (2018). Negative body image and cognitive biases to body size. (Doctoral dissertation). Retrieved from: https://hydra.hull.ac.uk/resources/hull:16444

Tovée, M. J., Benson, P. J., Emery, J. L., Mason, S. M., & Cohen-Tovée, E. M. (2010). Measurement of body size and shape perception in eating-disordered and control observers using body-shape software. *British Journal of Psychology, 94*, 501-516. doi:10.1348/000712603322503060

Tovée, M. J., & Cornelissen, P. L. (1999). The mystery of female beauty. *Nature, 399*, 215–216. doi:10.1038/20345

Tovée, M. J., & Cornelissen, P. L. (2001). Female and male perceptions of female physical attractiveness in front-view and profile. *British Journal of Psychology, 92*, 391–402. doi:10.1348/000712601162257

Tovée, M. J., Hancock, P. J. B., Mahmoodi, S., Singleton, B. R. R., & Cornelissen, P. L. (2002). Human female attractiveness: Waveform analysis of body shape. *Proceedings of the Royal Society of London B: Biological Sciences, 269*, 2205-2213. doi:10.1098/rspb.2002.2133

World Health Organisation (2003). Shaping the future report. Retrieved from http://www.who.int/whr/2003/en/

Yu, H., Garrod, O. G., & Schyns, P. G. (2012). Perception-driven facial expression synthesis. *Computers & Graphics, 36*, 152–162. doi:10.1016/j.cag.2011.12.002

Figure Legends

Figure 1. Screenshots of the stimuli from two consecutive trials from: (a) Experiment 1 with large bubbles, and (b) Experiment 2 with small bubbles. The first two columns show the stimuli as presented to the participant. Columns three and four show the same images but with a red outline to indicate the outline of the female model in the stimulus, beneath the gray overlay. On every trial, participants are given a partial view of the female model through a set of so-called Gaussian bubbles. These are circular holes with blurred edges that perforate the gray overlay that covers the model in the stimulus image. Please note that much visual detail will be lost in this illustration, compared to the original stimuli displayed on a PC monitor.

Figure 2. Diagnostic images for (a) the big bubbles mask Experiment 1, top row, and (b) the small bubbles mask Experiment 2, bottom row. For the Accurate and Overestimate figures (left and middle columns), the white circles show the locations of bubbles where correct response rates were significantly above the 75% criterion based on the GLMMs. The heat maps represent the averaged and smoothed raw data that contributed to the GLMMs. For the Accurate – Overestimate figure (right column), the white circles show where the differences between the two groups of observers are significantly different from zero. The blue-cyan colours in the heat map show where over-estimators made more correct responses than accurate estimators. The red-yellow colours in the heat map show where accurate estimators made more correct responses than over-estimators.

Figure 3: Fixation density maps for accurate and overestimators across the four eye-tracking conditions. Each image represents the same stimulus model with a semi-transparent coloured overlay to indicate fixation density, reported in *z*-scores. The higher the *z*-score (from gray,

through green and yellow to red), the more time participants spent looking at a particular region on the body. Black contours represent 3*SD*s, within which most fixations lie.

Figure 4: Shows predicted marginal means together with their 95%CIs, for co-registered eye fixation and psychophysical data, across a set of horizontal slices. The locations on the model's body of the horizontal slices through the combined datasets are indicated by letter groups: *A*, *B*, & *C* and *D*, *E*, & *F*, for the large and small bubble mask datasets, respectively. For accurate estimators, solid black lines with black circles represent the eye-movement data and solid white lines with white circles the psychophysical data. For overestimators, solid black lines with black triangles represent the eye-movement data and solid white lines with white triangles the psychophysical data. There is a small horizontal offset in the x-axes between the eye-movement and psychophysical data, so that error bars do not overlap.

Table 1. Experiment 1 with large bubble masks: Participant characteristics

	Accurate (*n* =12)		Overestimate (*n* =12)		Accurate vs. Overestimate		
	M	*SD*	*M*	*SD*	*p*	*d*	*95%CI*
Participant characteristics							
Age (years)	23.67	5.65	22.25	4.37	.99	-0.28	(-1.12 – 0.56)
BMI (kg/m^2)	21.97	2.89	22.16	3.22	1.00	0.06	(-0.77 – 0.90)
Depression							
BDI score	15.17	9.84	17.75	11.28	.99	0.24	(-0.60 – 1.09)
Body shape and eating concerns							
BSQ-16 score	38.92	20.78	47.67	23.03	.93	0.40	(-0.45 – 1.25)
EDE-Q global score	1.32	1.02	2.23	1.57	.50	0.69	(-0.48 – 1.55)
EDE-Q res score	1.40	1.37	2.03	1.32	.86	0.47	(-0.38 – 1.32)
EDE-Q eat score	0.45	0.52	1.28	1.42	.35	0.78	(-0.09 – 1.65)
EDE-Q wc score	1.58	1.45	2.47	1.84	.76	0.53	(-0.32 – 1.39)
EDE-Q sc score	1.86	1.67	3.14	2.17	.55	0.66	(-0.20 – 1.52)
Psychophysical performance							
PSE (kg/m^2)	22.16	2.98	25.85	3.40	.07	1.15	(0.25 – 2.06)
DL (kg/m^2)	0.67	0.26	1.15	0.88	.43	0.73	(-0.14 – 1.59)
Overestimation (PSE - BMI)	0.19	0.78	3.69	1.31	< .001	3.25	(1.98 – 4.53)
Mean bubble count	5.00	1.08	5.12	1.21	.92	0.15	(-0.69 – 0.99)
Mean percentage trials correct	74.41	0.86	74.33	1.10	.97	-0.08	(-0.92 – 0.76)

Note. BDI = Beck Depression Inventory; BSQ-16 = Body Shape Questionnaire; EDE-Q = Eating Disorders Examination Questionnaire global score; EDE-Q subscales: res = restraint; eat = eating concerns; wc = weight concerns; sc = shape concerns.

Table 2. Experiment 2 with small bubble masks: Participant characteristics

	Accurate (*n* = 12)		Overestimate (*n* = 12)		Accurate vs. Overestimate		
	M	*SD*	*M*	*SD*	*p*	*d*	*95%CI*
Participant characteristics							
Age (years)	22.58	6.40	20.92	3.75	.98	-0.32	(-1.16 – 0.53)
BMI (kg/m^2)	22.55	4.80	21.99	2.94	0.99	-0.14	(-0.98 – 0.70)
Depression							
BDI score	9.50	8.60	16.83	11.34	.45	0.73	(-0.14 – 1.59)
Body shape and eating concerns							
BSQ-16 score	45.67	23.91	53.17	18.41	.97	0.35	(-0.49 – 1.20)
EDE-Q global score	1.74	1.66	2.65	1.29	.64	0.62	(-0.24 – 1.48)
EDE-Q res score	1.35	1.43	2.22	1.43	.65	0.61	(-0.25 – 1.46)
EDE-Q eat score	1.08	1.48	1.63	1.40	.95	0.38	(-0.46 – 1.23)
EDE-Q wc score	2.08	2.09	2.97	1.63	.85	0.47	(-0.38 – 1.32)
EDE-Q sc score	2.43	1.92	3.79	1.39	.33	0.81	(-0.06 – 1.68)
Psychophysical performance							
PSE (kg/m^2)	22.40	4.68	25.96	2.72	.20	0.93	(0.05 – 1.81)
DL (kg/m^2)	0.75	0.30	1.03	0.23	.11	1.04	(0.15 – 1.94)
Overestimation (PSE – BMI)	-0.15	0.57	3.97	1.35	<.001	3.98	(2.53 – 5.42)
Mean bubble count	18.42	5.51	19.62	3.86	.74	0.25	(-0.59 – 1.09)
Mean percentage trials correct	74.75	1.46	74.03	1.05	.61	-0.33	(-1.17 – 0.51)

Note. BDI = Beck Depression Inventory; BSQ-16 = Body Shape Questionnaire; EDE-Q = Eating Disorders Examination Questionnaire global score; EDE-Q subscales: res = restraint; eat = eating concerns; wc = weight concerns; sc = shape concerns.

Table 3. Experiment 3: Participant characteristics

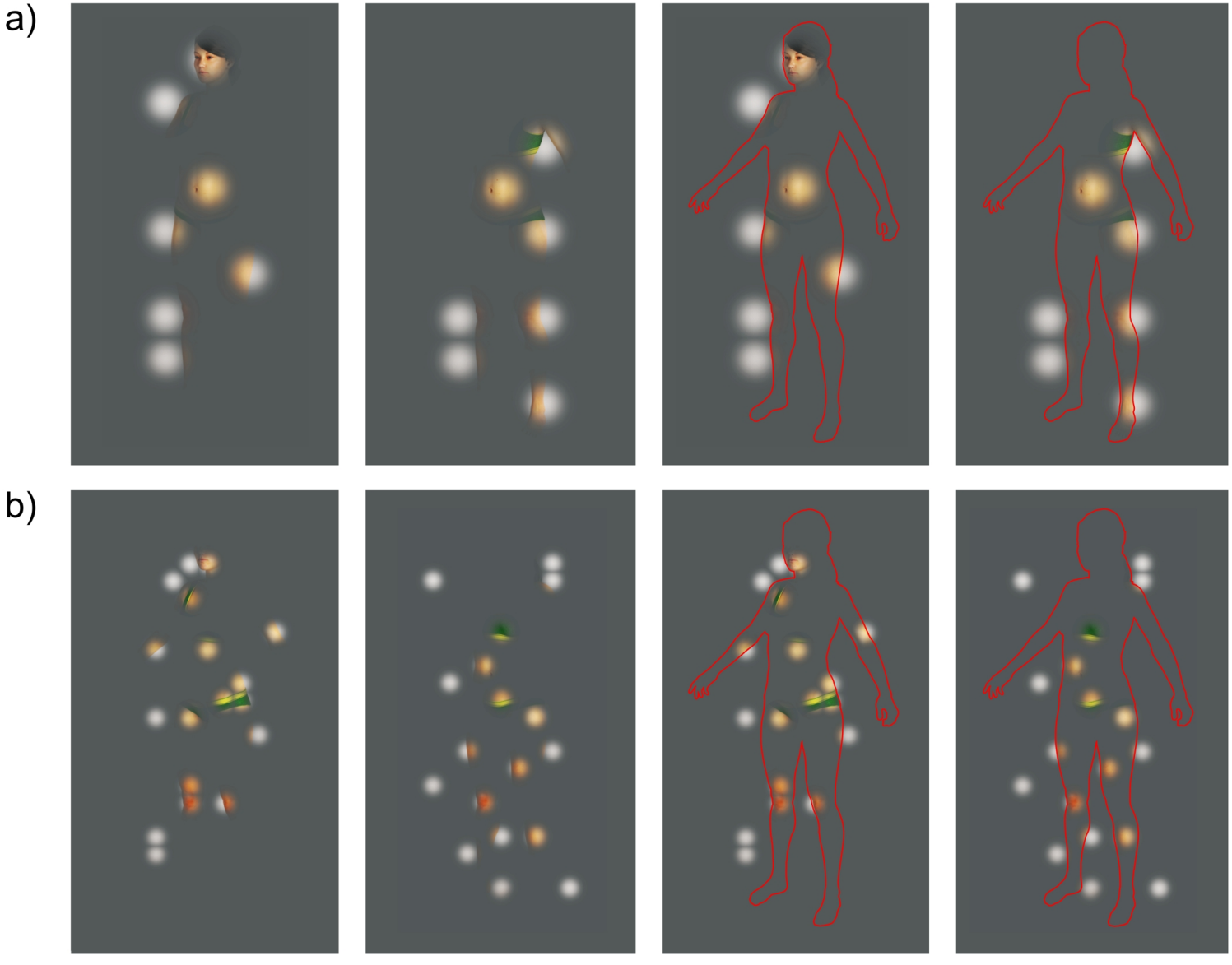
	Accurate (*n* = 12)		Overestimate (*n* = 12)		Accurate vs. Overestimate		
	M	*SD*	*M*	*SD*	*p*	*d*	*95%CI*
Participant characteristics							
Age (years)	20.58	1.98	23.33	5.35	.55	0.68	(-0.18 – 1.54)
BMI (kg/m^2)	23.42	6.38	21.73	3.57	.98	-0.33	(-1.17 – 0.52)
Depression							
BDI score	12.50	8.12	15.17	8.12	.98	0.33	(-0.52 – 1.17)
Body shape and eating concerns							
BSQ-16 score	38.92	16.81	46.33	12.61	.83	0.50	(-0.35 – 1.35)
EDE-Q global score	1.38	0.97	2.15	1.24	.53	0.69	(-0.35 – 1.35)
EDE-Q res score	1.15	0.92	2.37	1.89	.33	0.82	(-0.05 – 1.69)
EDE-Q eat score	0.65	0.69	1.12	0.96	.75	0.56	(-0.29 – 1.41)
EDE-Q wc score	1.55	1.48	2.32	1.11	.71	0.58	(-0.27 – 1.44)
EDE-Q sc score	2.16	1.55	2.79	1.35	.91	0.44	(-0.41 – 1.28)
Psychophysical performance							
PSE (kg/m^2)	23.45	6.15	25.35	3.27	.95	0.39	(-0.46 – 1.23)
DL (kg/m^2)	0.78	0.40	0.92	0.65	.99	0.25	(-0.59 – 1.09)
Overestimation (PSE – BMI)	0.03	0.64	3.62	2.03	< .001	2.39	(1.29 – 3.49)

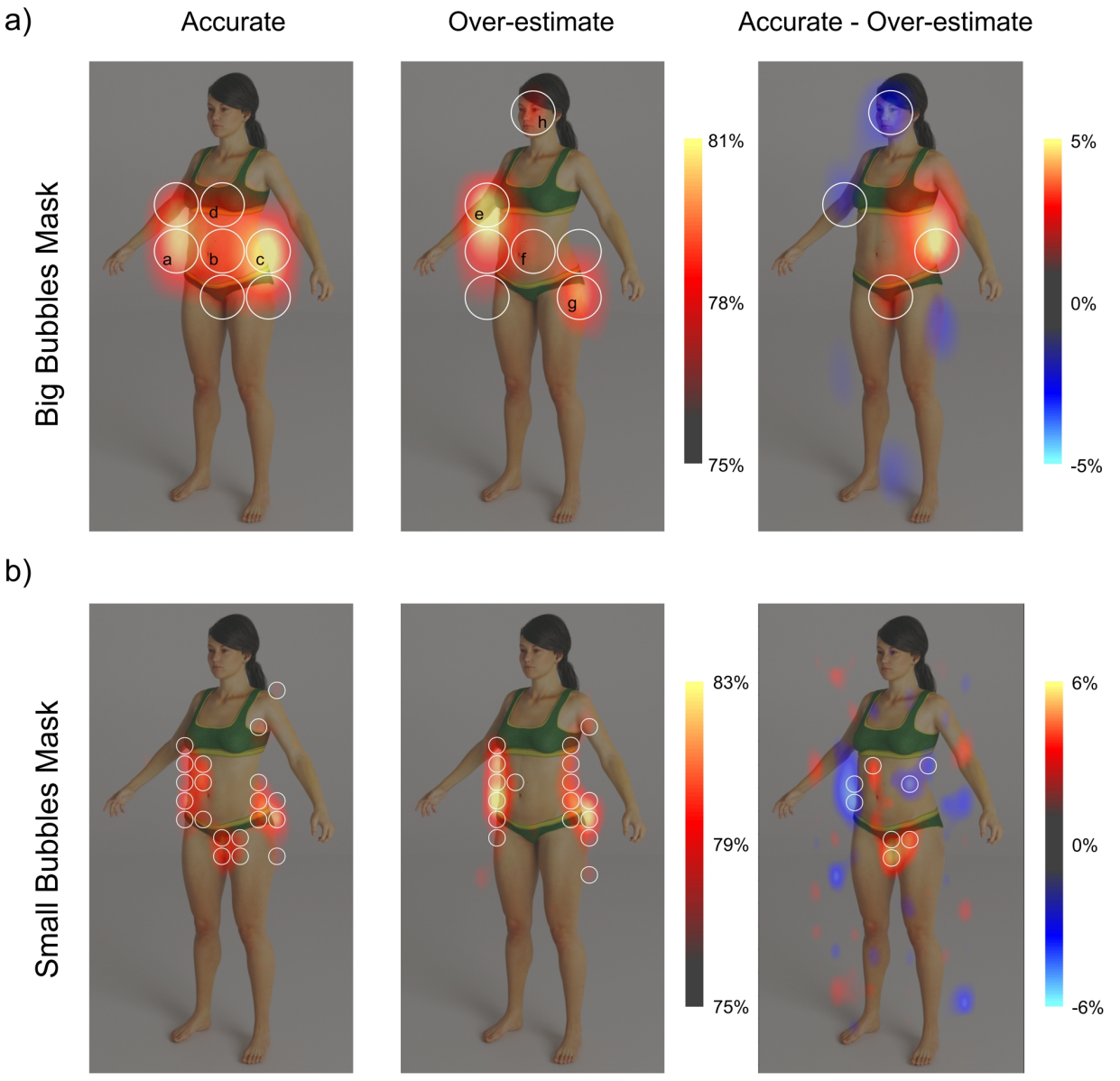
Note. BDI = Beck Depression Inventory; BSQ-16 = Body Shape Questionnaire; EDE-Q = Eating Disorders Examination Questionnaire global score; EDE-Q subscales: res = restraint; eat = eating concerns; wc = weight concerns; sc = shape concerns.

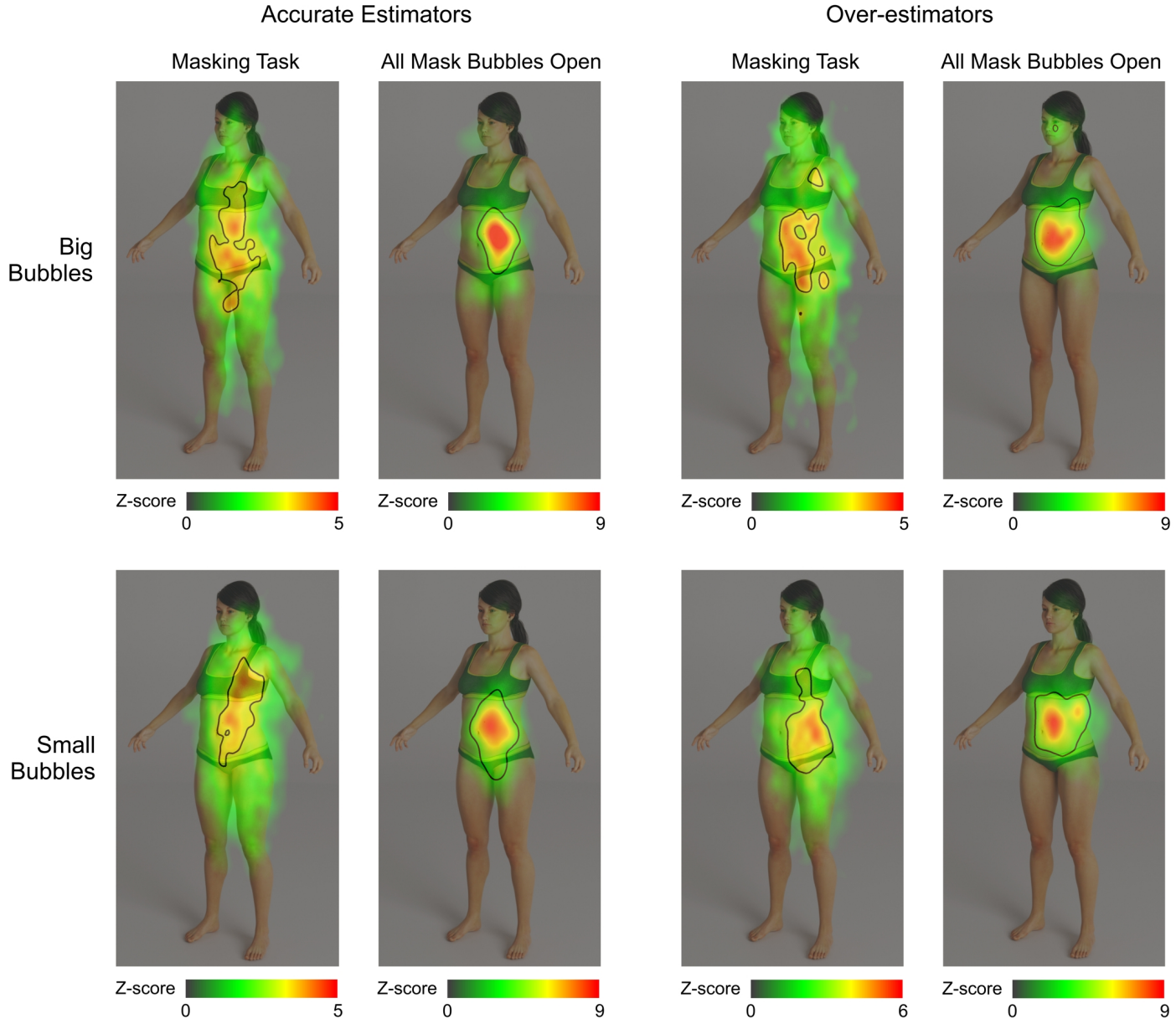
Table 4. Comparison of fixation density in each of the three columns.

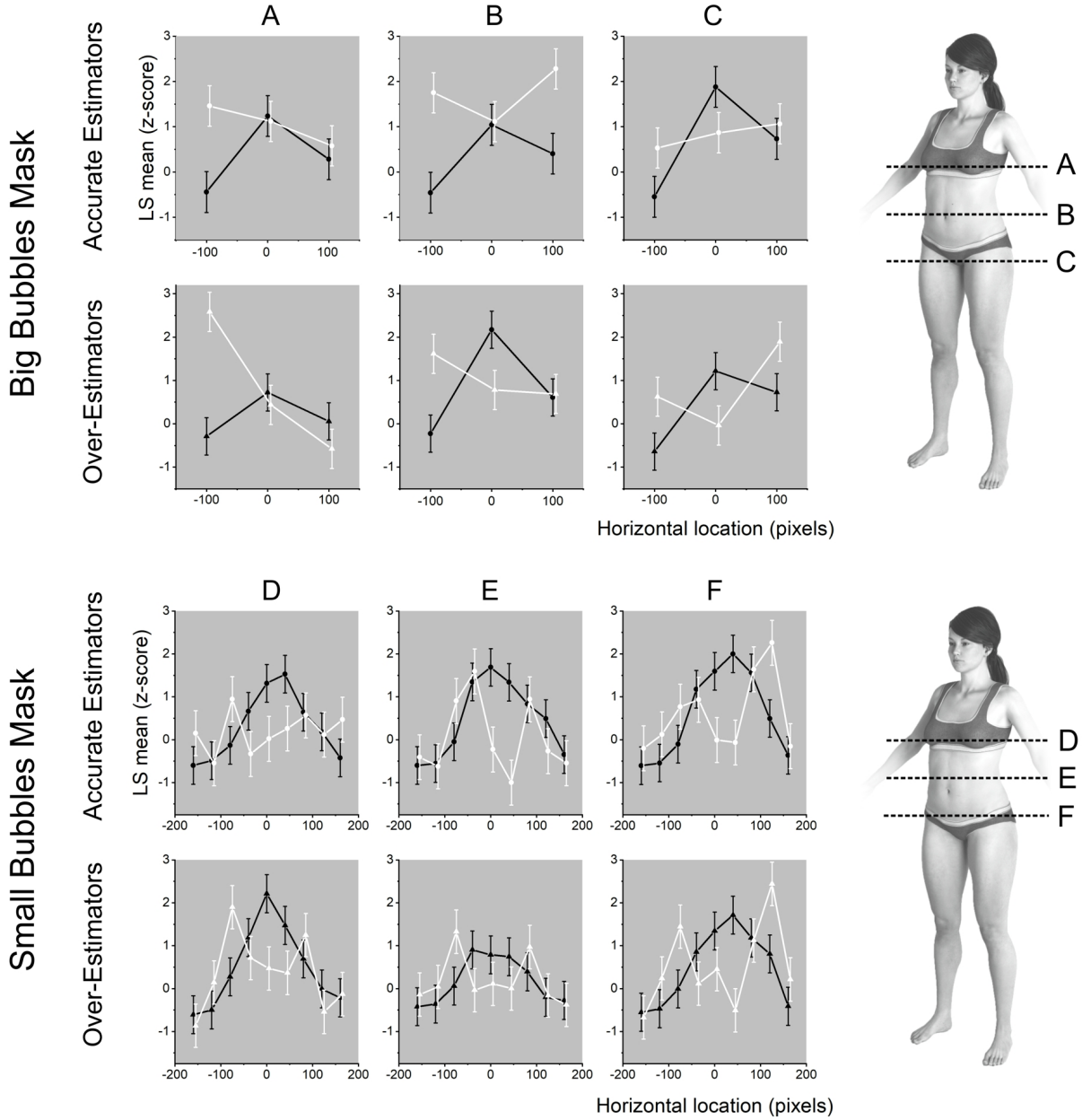
Group	Bubble Size	Task	Left column (%)		Middle column (%)		Right column (%)		Left vs. Middle	Right vs. Middle
			M	*SE*	*M*	*SE*	*M*	*SE*	*p*	*p*
Accurate	Big	Mask	32.41	4.66	56.98	3.74	10.60	2.26	< .001	< .001
		No mask	20.56	5.99	75.21	6.18	5.63	3.20	< .001	< .001
	Small	Mask	41.17	5.19	50.75	3.58	8.82	2.39	.08	< .001
		No mask	17.89	4.74	77.45	4.66	5.08	2.14	< .001	< .001
Over-estimate	Big	Mask	35.09	5.26	52.80	3.44	12.11	3.48	< .001	.003
		No mask	25.53	7.57	70.98	6.99	4.19	2.53	< .001	< .001
	Small	Mask	39.64	4.86	50.82	3.19	9.55	3.29	.05	< .001
		No mask	31.34	9.76	63.18	8.71	6.58	2.92	< .001	< .001

1084


a)
b)


a)
Accurate
Over-estimate
Accurate - Over-estimate
Big Bubbles Mask
a
b
c
d
e
f
g
h
81%
78%
75%
5%
0%
-5%
b)
Small Bubbles Mask
83%
79%
75%
6%
0%
-6%


Accurate Estimators
Over-estimators
Masking Task
All Mask Bubbles Open
Masking Task
All Mask Bubbles Open
Big Bubbles
Small Bubbles
Z-score
0
5
Z-score
0
9
Z-score
0
5
Z-score
0
9
Z-score
0
5
Z-score
0
9
Z-score
0
6
Z-score
0
9


Big Bubbles Mask
Small Bubbles Mask
Accurate Estimators
Over-Estimators
LS mean (z-score)
A
B
C
D
E
F
Horizontal location (pixels)